Transcrição

Continuando com as Medidas de Tendência Central, abordaremos a **mediana**.

Esta medida consiste no valor que **divide** a nossa série exatamente **ao meio**. Por exemplo, se tivermos as idades de dez pessoas, primeiro deveremos
ou seja, entre cinco mais novos e os cinco mais velhos.

Na parte "3.2 Mediana" de nosso notebook, teremos um roteiro de como obter a mediana com fórmulas específicas.

O primeiro passo é ordenar o conjunto de dados. Depois, identificaremos o número `n` de observações ou registros deste conjunto. Quando formos s

Quando "n" for **ímpar**, a posição do elemento mediano será obtido com:

$$Elemento_{Md} = \frac{n+}{2}$$

Quando "n" for **par**, será com:

$$Elemento_{Md} = \frac{n}{2}$$

Faremos somente o passo-a-passo do **ímpar** agora, mas recomendamos calcular com o "n" par como exercício extra, caso queira. Mais adiante, verei

Em nosso primeiro exemplo, teremos uma série desordenada com cinco valores: `[ 6, 4, 3, 9, 1]` . Quando organizarmos de forma crescente, obt

Como a quantidade de registros no conjunto é ímpar, aplicaremos sua fórmula correta. O "n" é este número "5" de observações somado com "1", resu

Este valor significará a terceira posição do conjunto, a qual é relativa ao **elemento mediano**, ou seja, o `4` será a medida mediana "Md" deste exercíc

Entenderemos esse cálculo na prática com nosso dataset `df` com o boletim de notas criado anteriormente como exemplo.

Na primeira célula desta parte do nosso notebook, criaremos a variável `notas_fulano` para o primeiro aluno. Esta será igual a `df.Fulano` e, em se

```
notas_fulano = df.Fulano
notas_fulano
```

Matemática	8
Português	10
Inglês	4
Geografia	8
História	6
Física	10
Química	8

Na célula seguinte, `notas_fulano` será igual a `notas_fulano` com o método Pandas **`.sort_value()`** que organizará os valores de maneira cresce

```
notas_fulano = notas_fulano.sort_values()
notas_fulano
```

Inglês	4
História	6
Matemática	8
Geografia	8
Química	8
Português	10
Física	10

Isso não precisará ser feito todas as vezes que quisermos calcular a mediana, mas é importante fazermos este passo-a-passo para a operação ficar ben

Logo, para entendermos bem a **posição mediana**, escreveremos `notas_fulano` com um outro método Pandas chamado **`.reset_index()`** que retira variável `index` do `DataFrame` .

```
notas_fulano = notas_fulano.sort_values()
notas_fulano
```

	index	Fulano
0	Inglês	4
1	História	6
2	Matemática	8
3	Geografia	8
4	Química	8
5	Português	10
6	Física	10

Com este novo índice numérico de `0` a `6` , veremos com clareza a posição `3` do elemento mediano `Geografia` com nota de Fulano `8` , pois temo

Para confirmarmos na prática, aplicaremos a fórmula na célula seguinte; escreveremos `n` sendo igual a `notas-fulano` com `.shape[]` recebendo

```
n = notas_fulano.shape[0]
n
```

O resultado da operação será `7`, correspondendo à quantidade de sete observações no conjunto de dados.

Na célula seguinte, obteremos `elemento_md` sendo igual a `(n + 1)` dividido por `2`. Depois, exibiremos `elemento_md`.

```
elemento_md = (n + 1) / 2
elemento_md
```

O elemento mediano apresentado está na posição `4.0`. De acordo com a tabela com `index`, corresponde à `Geografia` identificada pelo número `3`

Como se trata de um conjunto de dados bastante pequeno, este cálculo está bem simples. Mas em muitas situações, teremos uma grande tabela para l

É importante nos atentarmos às diferenças de cálculo entre quantidades pares e ímpares, conforme as fórmulas apresentadas.

Continuando, escreveremos `notas_fulano` na célula seguinte, e aplicaremos o **`loc[]`** que selecionará um item dentro do `DataFrame` ou *series*.

Dentro, passaremos o elemento mediano `elemento_md` com **`-1`** para encontrá-lo. É importante não esquecermos desta subtração.

```
notas_fulano.loc[elemento_md - 1]
```

O resultado do `index` será `Geografia` com a nota `8` de `Fulano`, cuja identificação é `3`.

Para calcularmos o mesmo valor utilizando a biblioteca Pandas para não precisarmos fazer tantas contas, chamaremos apenas `notas_fulano` com o

```
notas_fulano.median()
```

O retorno exibirá a mediana correspondente à nota `8.0` de `Fulano`.

Seguindo no notebook, encontraremos o "Exemplo 2" relativo à quantidade par de registros. Para auxiliar no experimento, pegaremos as notas de `Be` apenas `6` registros por meio de `sample()`.

O parâmetro `random_state` é um gerador de número aleatório. Repetindo o valor `101` apresentado aqui, será possível obter exatamente a mesma s

```
notas_beltrano = df.Beltrano.sample(6, random_state = 101)
notas_beltrano
```

Matemática	8
Inglês	4
Física	10
História	6
Química	8
Português	10

Realizando toda a operação necessária para encontrar a mediana, obteremos o valor `6.25` de `notas_beltrano` ao final do exercício com `n` sendo

```
notas_beltrano.median()
```

Para realizarmos a análise descritiva dos dados de nosso projeto, iremos para a parte "Obtendo a mediana em nosso dataset".

Na primeira célula, começaremos pela `Renda` . Recomendamos que também faça com as demais variáveis como exercício.

```
dados.Renda.median()
```

O resultado da operação será `1200.0` .

Há uma outra maneira de obtermos o resultado que será melhor abordada adiante nas Medidas Separatrizes; com o método **`quantile()`** vazio, teren

Isso acontece porque seu default é `q=0.5` , o que significa que partiremos o conjunto exatamente na metade para pegarmos a mediana da mesma forı apenas `?` no lugar dos parênteses do método e executando a célula.

```
dados.Renda.quantile?
```

Portanto, também é possível obtermos a mediana com `quantile()` , bem como outros valores que serão vistos mais adiante em Medidas Separatrize

A seguir, falaremos da medida **moda**.